Deep Time Series Attention Models for Crop Yield Prediction and Insights

Tryambak Gangopadhyay
Department of Mechanical Engineering
Iowa State University
Ames, IA 50011
tryambak@iastate.edu

Johnathon Shook
Department of Agronomy
Iowa State University
Ames, IA 50011
jmshook@iastate.edu

Asheesh K. Singh
Department of Agronomy
Iowa State University
Ames, IA 50011
singhak@iastate.edu

Soumik Sarkar
Department of Mechanical Engineering
Iowa State University
Ames, IA 50011
soumiks@iastate.edu

Abstract

Soybean yield depends on both environmental and genetic factors which can be difficult to dissect as multiple replications of many genotypes are needed in diverse environmental conditions. We aim to develop a prediction-based framework with multivariate time series as input comprising different weather variables, maturity groups, and genotype information. Based on historical performance records from Uniform Soybean Tests (UST) in North America, we present a deep learning model to analyze how key weather events and genetic interactions can affect yield. We use our proposed temporal attention model for a better understanding of the important time steps in predicting annual yield. We explain the learned attention weights from a domain knowledge perspective. We compare the results with that of the baseline model (using only stacked LSTMs) to ensure that we don't compromise on the performance while improving the understanding.

1 Introduction

Predicting crop yield is one of the significant challenges to address the food security issue. Accurate prediction of yield well before harvest along with useful insights can help in preventing famine and improving agricultural monitoring across diverse climatic conditions. Soybean has a long history of cultivation in North America, with the first reported production in Georgia in 1766 [1]. Over the years, production has expanded as far west as Kansas-Colorado border and has grown from southern Texas to Canada [2, 3]. Climatic resiliency is an important objective for plant breeders and farmers to get a high yield. The climatic variability can be associated with changes in temperature and rainfall events (including patterns and magnitude). In addition to spatial variability, temporal variability of weather variables [4] is equally important and generally less understood or not included in yield prediction studies. Prediction of the effects of changing environments on performance can help in making informed marketing decisions, optimizing production and comparing results over multiple years [5].

North American annual soybean yield trials (known as Uniform Soybean Tests (UST)) have been coordinated through the United States Department of Agriculture (USDA) since 1941 [6, 7]. These trials are used to evaluate current and pre-commercial varieties in multiple environments within their range of adaptation. Traditionally, crop growth models have been proposed to predict yield [8], but

Second Workshop on Machine Learning and the Physical Sciences (NeurIPS 2019), Vancouver, Canada.

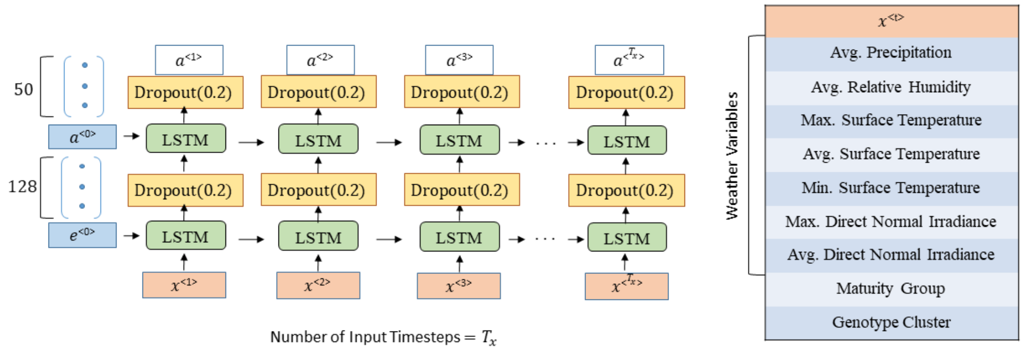

Figure 1: LSTMs are used for encoding the input sequence which is of length T_x and the output from the first LSTM layer is a batch of sequences which are propagated through another layer of LSTM. We use dropout regularization after each LSTM layer to prevent overfitting.

these models may have deficiencies related to estimating input parameters and predicting in complex, unforeseen circumstances [9]. Traditional linear methods such as AutoRegressive Integrated Moving Average (ARIMA) mainly focus on univariate data and may not be efficient in multivariate problems. Deep learning models can provide solutions in the presence of such complex data comprising of different weather variables, maturity groups, and genotype information.

For time series prediction tasks, deep neural networks show robustness to noisy inputs and also have the capability to approximate arbitrary non-linear functions [10]. Long Short Term Memory (LSTM) Networks [11] are very useful for time series analysis as they can capture the long-term temporal dependencies in complex multivariate sequences [12]. LSTMs have shown state-of-the-art results in various applications including off-line handwriting recognition [13] and natural language processing. LSTMs have also been used effectively for multivariate time series prediction tasks [14, 15, 16]. Previous work [17] using deep learning for yield prediction has utilized multi-spectral images to predict yield (instead of leveraging only multivariate time series an input) without model interpretability.

We propose a framework based on stacked LSTMs and temporal attention to predict the yearly value of crop yield. The input is a multivariate time series comprising of the entire crop season (the US and Canada UST data for years 2003-2015). We vary the number of input time steps and compare the performance of our proposed model with the baseline model for two variations of each model. The temporal attention mechanism highlights the significant periods throughout the year leading to high or low yield prediction, which we validate using domain knowledge.

2 Experiments

2.1 Models

We demonstrate the performance of our proposed model (Temporal Attention Model) along with the baseline model without using attention (Stacked LSTMs Model). This is a many-to-one prediction problem and the output of both models is yearly yield. We have two variants for each model depending on using the maturity group and genotype cluster information or not.

2.1.1 Stacked LSTMs Model

LSTM networks comprise of input, output and forget gates which helps in preventing the memory contents from being perturbed by irrelevant information. For encoding the input sequence we use two stacked LSTM layers to get the T_x annotations as shown in Fig. 1. In the Stacked LSTMs Model, the last hidden state of the encoding part is assumed to be the compressed representation from the entire input sequence. This fixed-dimensional representation is used for predicting the output value of yield (Fig. 2).

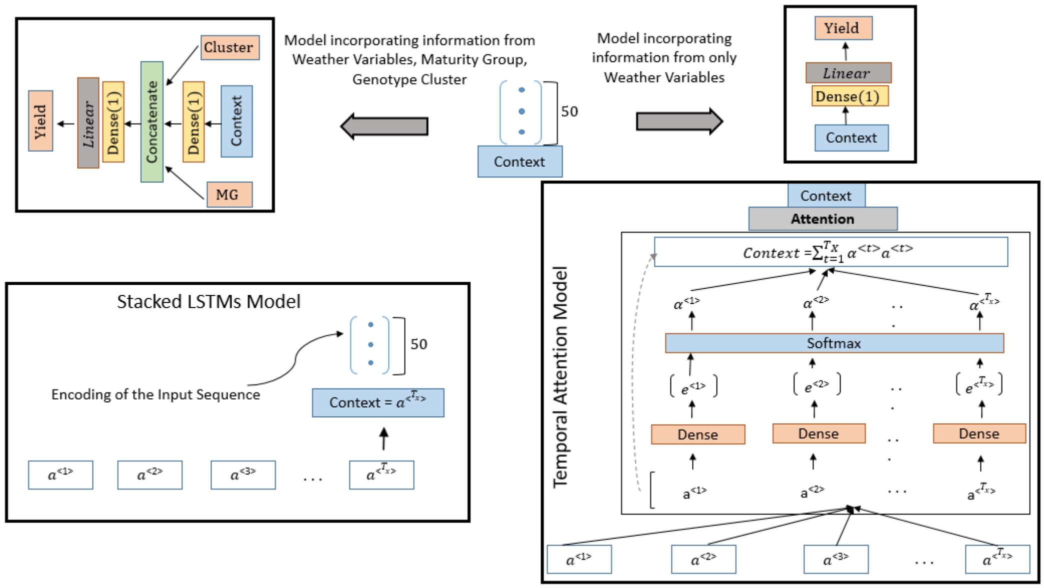

Figure 2: Temporal Attention Model takes in a sequence of annotations as input and generates the context vector. The Stacked LSTMs model is shown along with the two variants of each model based on the input information.

T_x	Model	RMSE(Weather Variables)	RMSE(Including all)
7	Temporal Attention	8.271	7.136
	Stacked LSTMs	8.246	7.216
15	Temporal Attention	8.253	7.128
	Stacked LSTMs	8.251	7.205
30	Temporal Attention	8.236	7.148
	Stacked LSTMs	8.288	7.150

Table 1: Comparison of performance of the two models using test set RMSE values (there are 2 variants of each model based on the input information) with varying input sequence length (T_x). The input of one variant includes only weather variables and the other variant has all inputs (weather variables, MG and cluster).

2.1.2 Temporal Attention Model

Attention mechanism can soft search the most relevant input timesteps [18] improving the performance of the Encoder-Decoder Model [19, 20]. Temporal attention can be applied in the context of many-to-many time series prediction problems [16]. For this many-to-one prediction problem, we implement the attention mechanism as shown in Fig. 2. Our approach doesn't involve using a decoder LSTM as we are predicting only one value for a particular input sequence and thus, the attention block doesn't use the hidden state of the post-attention LSTM. The input sequence is encoded into a sequence of vectors (annotations) instead of encoding into a fixed-length vector. The context vector is generated by taking a weighted sum of the annotations. The attention weights for each annotation is computed using softmax. The alignment model scores how well the inputs around timestep t is aligned with the prediction. The training time and the number of trainable parameters of this model remain almost the same as that of the Stacked LSTMs model.

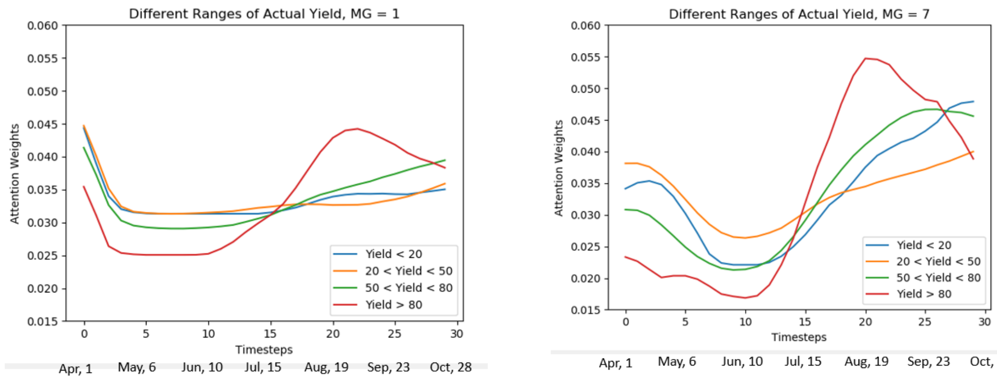

Figure 3: Distribution of attention weights for the entire input sequence (spanning the growing season). The results are demonstrated for two different maturity groups (MG = 1, MG = 7) considering ranges of actual yield.

2.2 Dataset - Preparation of Performance Records, Acquisition of Weather Records & Genotype Clustering

Files from 2003-2015 are downloaded as PDFs [6, 7]. The tables are manually curated to align all performance records for corresponding genotype/location combinations. Records which do not have yield data (due to a variety not being planted in a specific location, or dying prior to production of seed) are not considered. The final dataset comprises a total of 104320 performance records over the 13 year period representing 5609 uniques genotypes. The original PDFs, including methods used for all measurements, can be found here [6] for the North, and here [7] for the South. Daily weather records for all location/year combinations are compiled based on the nearest weather station available (25km grid) from Weather.com. We downsample the dataset to include maximum, minimum, and average conditions on different time frames throughout the growing season (defined April 1st through October 31) and we append this information to the performance records.

We incorporate the information of maturity group zones which are determined by abiotic factors. We include the genotype-specific criteria to apply the model for specific genotypes rather than mean location yield across genotypes. We develop a completely connected pedigree for all lines with available parentage information, resulting in the formation of a 5609 x 5609 correlation matrix. We cluster the genotypes in 5 clusters using the K-means Clustering technique based on this correlation matrix to extract information about relatedness. With this hard clustering technique, each genotype belongs to any one of the 5 clusters.

3 Results

Both the models have RMSE of around 8.2 bushels/acre when only weather variables are incorporated in the input sequence. We achieve improved performance by including the maturity group and genotype cluster information as shown in Table 1. The RMSE values then drop to around 7.1-7.2 bushels/acre. The results of both the models are comparable for different input sequence length (T_x). This confirms that we are not sacrificing performance by implementing temporal attention.

The plots in Fig. 3 shed light on the patterns followed across genotypes, when grouped in four groups: seed yield in bushels/acre (<20, 20-50, 50-80, 80). While somewhat arbitrary, these groups allow more realistic trend comparisons of low and high performing genotype. Further, we investigate these trends in two geographically distinct maturity groups: MG1 (Northern US adaptation) and MG7 (Southern US adaptation). A most interesting observation is that early-season variables are less important for yield prediction in the highest yielding genotypes. The mild sigmoid curve is observed in both MG1 and MG7 for the highest yielding group. While these are based on prediction ability, it points to the increasing importance of features in the August – September time phases for both North and South US regions. These times coincide with crop reproductive phases, emphasizing their importance in the final yield. Since our results are built on more than a decade of data, it also reflects that early stages are less useful in seed yield prediction and the plasticity of soybean genotypes in earlier stages of growth and development. Interestingly, the remaining groups follow a somewhat similar trend across the crop season.

4 Conclusions

Accurate prediction of yearly yield can significantly impact the economies of agricultural states and daily lives through food prices. We leverage a large amount of data based on historical records to train our model and achieve accurate predictions. Our proposed framework can be useful in different applications like agriculture production (yield estimation and crop selling decisions), plant breeding (making effective selections) and field experimentation (selecting productive sites). As explained by domain knowledge, the interpretations provided by the model can aid in making more informed decisions to investigate other crop species as well. The future work includes the development of a model that can provide more insights apart from identifying the important periods of the crop growing season.

References

[1] Theodore Hymowitz and Jack Rodney Harlan. Introduction of soybean to north america by samuel bowen in 1765. *Economic Botany*, 37(4):371–379, 1983.

[2] Soybeans: Planted acreage by county. `https://www.nass.usda.gov/Charts_and_Maps/Crops_County/sb-pl.php`. Accessed: 2018-01-16.

[3] Canadian soybean seeded acres (1980 to current). `http://soycanada.ca/statistics/seeded-area-acres/`. Accessed: 2018-01-16.

[4] Agriculture. `https://nca2014.globalchange.gov/report/sectors/agriculture`. Accessed: 2018-01-16.

[5] Shrikant S Jagtap and James W Jones. Adaptation and evaluation of the cropgro-soybean model to predict regional yield and production. *Agriculture, ecosystems & environment*, 93(1-3):73–85, 2002.

[6] Uniform soybean tests, northern region. `https://www.ars.usda.gov/midwest-area/west-lafayette-in/crop-production-and-pest-control-research/docs/uniform-soybean-tests-northern-region/`. Accessed: 2018-01-10.

[7] Uniform soybean tests. `https://www.ars.usda.gov/southeast-area/stoneville-ms/crop-genetics-research/docs/uniform-soybean-tests/`. Accessed: 2018-01-10.

[8] Élodie Blanc. Statistical emulators of maize, rice, soybean and wheat yields from global gridded crop models. *Agricultural and Forest Meteorology*, 236:145–161, 2017.

[9] Michael J Roberts, Noah O Braun, Thomas R Sinclair, David B Lobell, and Wolfram Schlenker. Comparing and combining process-based crop models and statistical models with some implications for climate change. *Environmental Research Letters*, 12(9):095010, 2017.

[10] Georg Dorffner. Neural networks for time series processing. In *Neural network world*. Citeseer, 1996.

[11] Sepp Hochreiter and Jürgen Schmidhuber. Long short-term memory. *Neural computation*, 9(8):1735–1780, 1997.

[12] Pankaj Malhotra, Lovekesh Vig, Gautam Shroff, and Puneet Agarwal. Long short term memory networks for anomaly detection in time series. In *Proceedings*, page 89. Presses universitaires de Louvain, 2015.

[13] Patrick Doetsch, Michal Kozielski, and Hermann Ney. Fast and robust training of recurrent neural networks for offline handwriting recognition. In *Frontiers in Handwriting Recognition (ICFHR), 2014 14th International Conference on*, pages 279–284. IEEE, 2014.

[14] Johnathon M Shook, Linjiang Wu, Tryambak Gangopadhyay, Baskar Ganapathysubramanian, Soumik Sarkar, and Asheesh K Singh. Integrating genotype and weather variables for soybean yield prediction using deep learning. *bioRxiv*, page 331561, 2018.

[15] Zehui Jiang, Chao Liu, Nathan P Hendricks, Baskar Ganapathysubramanian, Dermot J Hayes, and Soumik Sarkar. Predicting county level corn yields using deep long short term memory models. *arXiv preprint arXiv:1805.12044*, 2018.

[16] Tryambak Gangopadhyay, Sin Yong Tan, Genyi Huang, and Soumik Sarkar. Temporal attention and stacked lstms for multivariate time series prediction. 2018.

[17] Jiaxuan You, Xiaocheng Li, Melvin Low, David Lobell, and Stefano Ermon. Deep gaussian process for crop yield prediction based on remote sensing data. In *Thirty-First AAAI Conference on Artificial Intelligence*, 2017.

[18] Dzmitry Bahdanau, Kyunghyun Cho, and Yoshua Bengio. Neural machine translation by jointly learning to align and translate. *arXiv preprint arXiv:1409.0473*, 2014.

[19] Kyunghyun Cho, Bart Van Merriënboer, Caglar Gulcehre, Dzmitry Bahdanau, Fethi Bougares, Holger Schwenk, and Yoshua Bengio. Learning phrase representations using rnn encoder-decoder for statistical machine translation. *arXiv preprint arXiv:1406.1078*, 2014.

[20] Ilya Sutskever, Oriol Vinyals, and Quoc V Le. Sequence to sequence learning with neural networks. In *Advances in neural information processing systems*, pages 3104–3112, 2014.